Factors That Affect Teaching Scores in Economics Instruction: Analysis of Student Evaluation of Teaching (SET) Data

Mohammad Alauddin* & Clem Tisdell, School of Economics Discussion Paper No. 353, December 2007,
School of Economics, The University of Queensland. Australia.

Full text available as:
PDF- Requires Adobe Acrobat Reader or other PDF viewer

Abstract

This paper explores the factors that affect students' evaluations of economics instructions using a sample of over 2400 completed questionnaires at a large Australian university. Ordered probit analysis is used to determine the changes in the predicted probability of teaching evaluation (TEVAL) scores with variations, amongst other things, in students' perceptions of the quality of presentation; explanation and organization of lecture material; and helping students improve their learning skills. Analyses of the comparative importance of the relationships both for undergraduate and postgraduate courses reveal significant differences across levels of the undergraduate program but little differences in students' responses in higher level undergraduate and postgraduate instructions. One disturbing finding is that a key variable, namely emphasis on thinking rather than memorizing (THINKMEM) has little or no substantive impact on TEVAL. Thus the implication is that high TEVALs can be achieved at the cost of some critically important factors in teaching and learning. Consequently, those using just TEVAL score to evaluate teaching need to look closely at other factors of critical importance.

EPrint Type:	Departmental Technical Report
Keywords:	Economics teaching, predicted probability, student teaching evaluation, teaching effectiveness, university education
Subjects:	340000 Economics;
ID Code:	JEL Classification A2, I2
Deposited By:	

Mohammad Alauddin
School of Economics, University of Queensland,
Brisbane, Queensland, Australia;
Phone: 3365 6664 Fax: 3365 7299
E-mail: m.alauddin@economics.uq.edu.au
*corresponding author

Factors That Affect Teaching Scores in Economics Instruction: Analysis of Student Evaluation of Teaching (SET) Data

Mohammad Alauddin* & Clem Tisdell
School of Economics
The University of Queensland
Brisbane, Qld. 4072 Australia
Fax: 61 7 3365 7299

Abstract

This paper explores the factors that affect students' evaluations of economics instructions using a sample of over 2400 completed questionnaires at a large Australian university. Ordered probit analysis is used to determine the changes in the predicted probability of teaching evaluation (*TEVAL*) scores with variations, amongst other things, in students' perceptions of the quality of presentation; explanation and organization of lecture material; and helping students improve their learning skills. Analyses of the comparative importance of the relationships both for undergraduate and postgraduate courses reveal significant differences across levels of the undergraduate program but little differences in students' responses in higher level undergraduate and postgraduate instructions. One disturbing finding is that a key variable, namely emphasis on thinking rather than memorizing (*THINKMEM*) has little or no substantive impact on *TEVAL*. Thus the implication is that high *TEVALs* can be achieved at the cost of some critically important factors in teaching and learning. Consequently, those using just *TEVAL* score to evaluate teaching need to look closely at other factors of critical importance.

Key words: Economics teaching, predicted probability, student teaching evaluation, teaching effectiveness, university education.
JEL Classification: A2, I2.

* Corresponding author: m.alauddin@uq.edu.au

Factors That Increase Scores in Economics Instruction: Analysis of Student Evaluation of Teaching (SET) Data

1 INTRODUCTION

Student evaluation of teaching is burgeoning. As Wilson (1998, p.A12) states "…Only about 30 per cent of colleges and universities asked students to evaluate professors in 1973, but it is hard to find an institution that doesn't today. And student ratings carry more and more weight. … Such evaluations are now the most important, and sometimes the sole, measure of an instructor's teaching ability". It is invariably used in promotion or tenure decisions as the most important indicator of teaching 'quality'. This notwithstanding, there is considerable controversy surrounding the derivation and use of teaching effectiveness instruments [1]. Despite their widespread use, student evaluations of teaching are inexact. As Mason, Steagall and Fabritius (1995, p.403) point out that:

> "…Students are not fully informed consumers because they do not necessarily know whether the professor is providing them with the relevant material, and doing so correctly. Consequently, students' judgment may be insufficiently well informed to evaluate this portion of the performance of their professors. Furthermore, students may not be fully cognizant of the quality until later experiences dictate the long-term value transferred.[2] In addition, the methodological approaches employed by a professor may be effective for a

[1] For an excellent summary of the controversy see, amongst others, Mason et al. (1995); Wilson (1998); Marsh (1987); Marsh and Roche (1997); Greenwald and Gilmore (1997); d'Appollonia and Abrami (1997); Mckeachie (1997); Becker (2000); Aleamoni (1999); Krautmann and Sander (1999); Wright (2006) and Becker (2000)

[2] Alauddin and Tisdell (2000, p.8) in expressing a similar view stated that ' … the quality of a program lies not necessarily in its immediate high approval rating but in appreciating the quality of value added in terms of analytical abilities of enduring character critical to a variety of situations encountered in a real world context. The real significance of this value added cannot conceivably be appreciated until well after one's completion of the degree and involvement in the workforce. … '. Consequences of asymmetry of information are well discussed in the literature since the pioneering work of Akerloff (1970).

particular student, or even the majority of students, but they are unlikely to the best for all of the students. Student evaluation scores will reflect both the views of those students for which (sic.) the methods works, and those for which (sic.) they do not …”.

This paper uses a large sample of student evaluation data on teaching (SET) and seeks answers to the following questions:

- What are the principal determinants of teaching effectiveness score?
- Do the impacts of these factors vary across postgraduate and undergraduate programs, and between levels within the undergraduate program?
- How does an increase or decrease in the score for any determinant affect the predicted probability of perceived teaching effectiveness?

In this paper Section 2 presents and outlines the main features of the data. Section 3 presents and discusses the empirical results. Section 4 provides and examines results of analysis of responsiveness of teaching effectiveness to instruction attributes. Section 5 presents a concluding overview and comments.

2. THE DATA: AN INTERPRETIVE ANALYSIS

The basic data for this study are from the SET surveys for nine economics courses that include four large second and two large third level undergraduate courses and three large postgraduate courses at the University of Queensland. These are for the period 2000 to 2006 and are based on 2467 completed SET forms. In this paper, we have regressed student responses on TEVAL scores. Of the 2467 responses, 1573 refer to the undergraduate samples across six courses at two levels while 894 relate to three postgraduate courses.

These are ‘official’ data. Note that these surveys do not include any factors that relate to student heterogeneity measured *inter alia* by study habits, attitude to learning, ethnicity, degree destinations, and English language competency. It would be desirable to pay greater attention to these factors in future studies[3]. The variable codes and definitions, dependent and independent variables and prior expectations about the direction of relationship with the dependent variable are provided in Table 1.

Table 1: Definitions of Variables and Description of SET Data

Variable Code	Description	Expected relation with TEVAL
TEVAL	**Dependent variable**: All things considered how would you rate this lecturer’s overall effectiveness as a university teacher? (1 – very poor, 5 – outstanding)	-
Independent variables: Instructor attributes ((1 – strongly disagree, 5 – strongly agree)		
ORGANIZE	The lecturer produced classes that were well organized	Positive
PRESENT	The lecturer presented material in an interesting way	Positive
FEEDBACK	The lecturer h=gave adequate feedback on my work	Positive
RESPECT	The lecturer treated students with respect	Positive
KNOWWELL	The lecturer seemed to know the subject well	Positive
ENTHUSM	The lecturer communicated his/her enthusiasm for the subject	Positive
THINKMEM	The lecturer emphasized thinking rather than memorizing	Positive
EXPLAIN	The lecturer gave explanations that were clear	Positive
CONSULT	The lecturer was available for consultation	Positive
LSKILLS	The lecturer helped to improve my learning skills	Positive

The data do not meet the criterion of strict randomness in the sense that courses could not be selected at random. This is because many staff members are sensitive to letting others use their *TEVAL* records for research. Nevertheless, the data used in this study relate to a large range of

[3] Mason et al (1995) and subsequently Sproule (2002) who included a range of variables to account for (i) instructor attributes; (ii) student attributes; and (iii) course attributes. Sproule (2002, p.289) went further in that he mathematically provided the proof for the underdetermination of instructor performance by SET data (see also Laudan and Leplin 1991). Wachtel (1998) provided a comprehensive list of background variables that influence the SET procedure. This list includes (i) characteristics associated with the administration of the SET; (ii) the characteristics of the instructor; (iii) the characteristics of the students; and (iv) the reaction to the dissemination and use and the SET. See also Felton et al. (2004).

courses – including large-sized second and third level undergraduate and postgraduate courses. These courses have a large degree of diversity in their student populations typified, amongst other things, by academic background, degree destination, and English language competency. Nevertheless, the fact remains that only the 'above average' instructors allowed us access to their *TEVAL* scores in this case.

Note also that the university requires all instructors to collect *TEVAL* data. The data collected represent the responses from only those students who are present in the class on the day *TEVAL* surveys are conducted. Thus, every student does not have an equal chance of appearing in the data. Those students who are less likely to attend classes are under-represented, a selection bias that most likely be of consequence. Since they may have chosen not to attend as frequently as others, they probably do so for a variety of reasons including 'lack of interest in lectures' and work or family commitments. This could result in skewed empirical results.

There is little that can be done about this except that we can analyze the responses that we actually receive from these students and acknowledge their limitations. Because of these limitations the results should be evaluated with caution. Nonetheless, the importance of our results should not be dismissed. For one thing, it is those students who attend and determine an instructor's *TEVAL* score.

Given the ordinal nature of the data, median and mode, not mean, are the appropriate measures of central tendency[4, 5]. The descriptive statistics reveal that the distributions of *TEVAL* (see Figure 1)

---

[4] The educational literature and the administrators alike routinely use the mean rather than median or mode even though it is patently wrong to do so from a statistical point of view in case of ordinal data. What seems intriguing is that some of the administrators are highly competent mathematicians, statisticians or econometricians who would advise their students to stick to methodological correctness when they teach. However, when wearing the administrators' hat such as head of school, or serving on promotion and tenure committees, they stridently defend the use of mean *TEVAL* score as the indicator of instructors' teaching quality. The heads or other administrators routinely express concern and give warning of failure to uphold (maintain) teaching quality if a staff member records a (mean) score of below 3.5 (on a five-point scale) in any course.

and other attributes are considerably skewed to the left implying a heavy concentration in the top end of the 5-point scale. In most cases, the highest point on the scale is in the third quartile (Q_3) while the first quartile (Q_1) without exception was located the 3-4 range.

For illustrative purposes, we tested to determine whether the distributions differed between postgraduate and undergraduate samples or between the two levels of the undergraduate program. This test was carried out only for *TEVAL*. Visual inspection of the relevant distributions presented in Figure 1 suggests that upper-level undergraduate and postgraduate samples have similar distributions while the lower and upper undergraduate distributions differ. A two-sample Kolmogorov-Smirnov test indicates that *TEVAL* distributions for the two undergraduate samples are significantly different[6]. The same test also reveals that the distributions for UG3 and the postgraduate samples do not differ significantly[7]. However, as expected the distributions relating the lower undergraduate sample and the postgraduate one are significantly different[8].

---

[5] The detailed descriptive statistics are not presented for brevity but are available upon request.

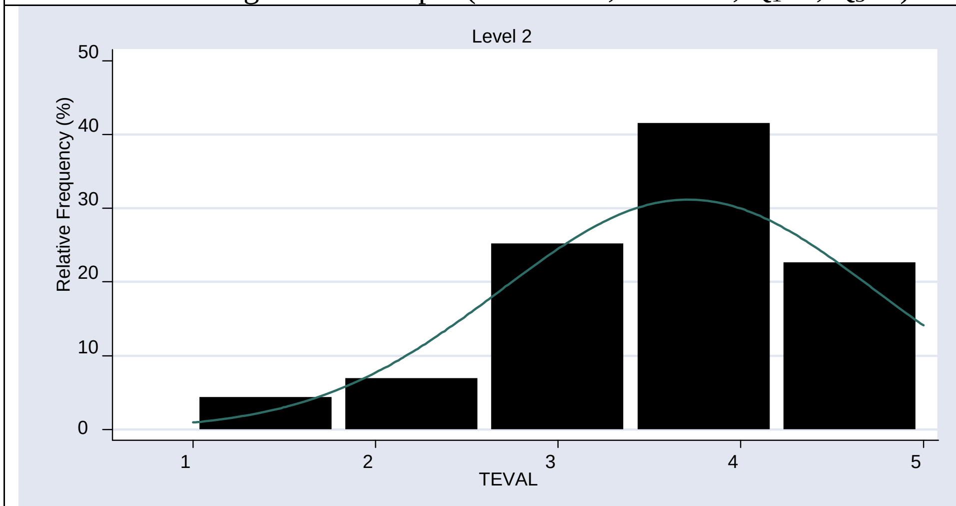
[6] The null hypothesis for a no difference between distributions was rejected (Kolmogorov-Smirnov *Z* statistic = 3.553, *p*-value =0.000).

[7] The null hypothesis for a no difference between distributions could not be rejected (Kolmogorov-Smirnov *Z* statistic = 0.431, *p*-value =0.992).

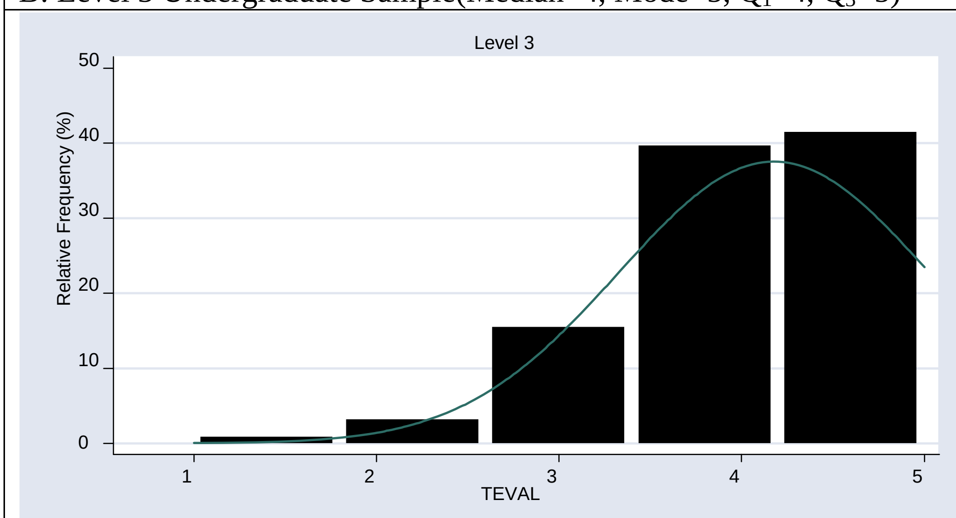
[8] The null hypothesis for a no difference between distributions was rejected (Kolmogorov-Smirnov *Z* statistic = 4.544, *p*-value =0.000).

Figure 1: Distributions of *TEVAL* for two levels of undergraduate and postgraduate samples

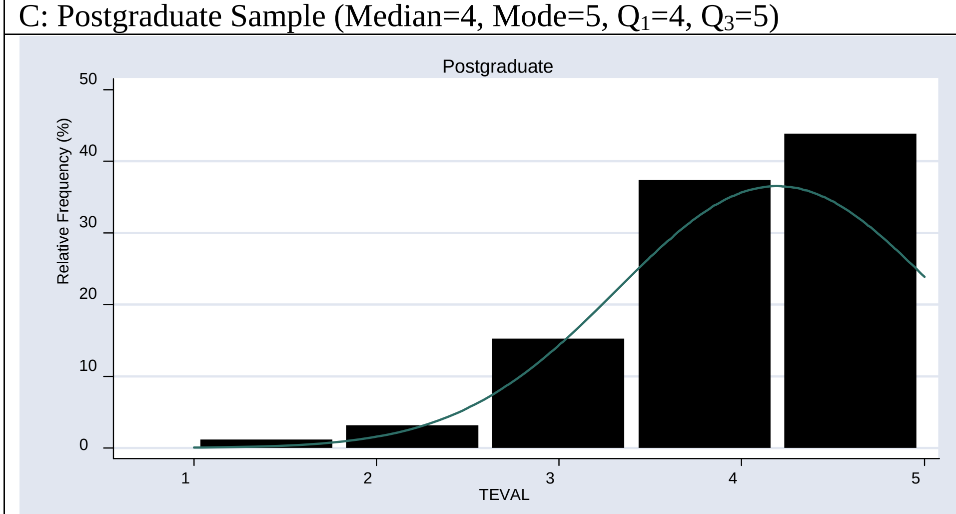
A: Level 2 Undergraduate Sample (Median=4, Mode=4, Q_1=3, Q_3=4)



B: Level 3 Undergraduate Sample(Median=4, Mode=5, Q_1=4, Q_3=5)



C: Postgraduate Sample (Median=4, Mode=5, Q_1=4, Q_3=5)



Note that a transition from the lower to the upper level undergraduate courses leads to decline in the spread of the distribution of *TEVAL* scores as one moves upwards. *Ceteris paribus* this indicates that those who teach lower level classes are likely to obtain lower and more dispersed *TEVAL* scores than those teaching higher level classes. However, at higher levels the differences between those distributions may not be statistically significant.

Several factors may explain this suggested pattern. For example, students in their earlier years may show considerable variation in cottoning on to a new subject. By later years, they are more familiar with its terminology and approach and may show less variation in their degree of learning about the subject. This may be reflected in their *TEVAL* scores. Furthermore, sorting is likely to occur. Those students who are less enthusiastic or less able to cope with a subject are less likely to continue with it in later years than those who are more capable and enthusiastic. This, in all probability, will be reflected in the distribution of the *TEVAL* scores. However, further research is warranted to identify the reasons for the observed changes in the distribution of *TEVAL* with the level of a subject. The results suggest that the *TEVAL* scores of those teaching lower level classes should be adjusted accordingly to be comparable with scores of those teaching higher level courses.[9]

[9] Pearson correlation coefficients between *TEVAL* and the remaining variables indicate significant correlations. However, it can also be seen that for the entire sample data *EXPLAIN, PRESENT, ORGANIZE and LSKILLS* show the strongest correlation with *TEVAL*. These results are similar to those of Tang (1997). These factors also seem to show similar strengths of correlation with *TEVAL* in the undergraduate and the postgraduate programs. The results are not presented here but are available upon request.

3 EMPIRICAL RESULTS

A large body of literature recognizes that linear regression is inappropriate when the dependent variable is categorical, especially if it is qualitative[10]. The appropriate theoretical model in such a situation is the ordered probit model (see for example, Greene 2000). For about three decades now these models have been widely used as a methodological framework for analyzing ordered data since the pioneering work of McKelvey and Zovoina (1975). In contrast to most of the education or educational psychology literature, the economics education literature uses ordered probit and/or multinomial logit models (DeCanio 1986; Mason et al. 1995; Boex 2000; Chan, Miller and Teha 2005).

The dependent variable, *TEVAL*, is coded from zero to four. Note that in terms of our model, a positive sign of any coefficient implies a higher probability of belonging to the highest category expressing ‘strong agreement’ or ‘best’. On the other hand, a negative sign implies a probability of probability of belonging to the lowest category ‘strong disagreement’ or ‘worst’.

Table 2 presents estimated equations for the two levels of undergraduate courses and one equation for the postgraduate sample using all the instruction attributes (as perceived by the students) listed in Table 1 as independent variables. The values of the pseudo-R^2 range between 0.38 and 0.51 indicating reasonable fits for all the models (Chan et al. 2005, p.30)[11]. An inspection of the results across levels and programs suggest that:

---

[10] Consider a customer survey where responses are coded 1 (worst/strongly disagree), 2, 3, 4 or 5 (best/strongly agree). ‘The linear regression model would treat the difference between a 4 and a 3 the same as that between a 3 and a 2, in fact they are only a ranking’ (Greene 2000, p.875).

[11] Since the traditional R^2 is poor measure of goodness of fit because even if a model fits perfectly R^2 will be less than one. Since the model is estimated using a maximum likelihood approach, a pseudo R^2 is defined by McFadden as R^2=1-(L_U/L_R). L_R is the restricted log likelihood, which is the value of the log of the likelihood function at iteration 0 where slope of all parameters are set to zero and L_U is the unrestricted log likelihood, which is the maximized value of log of the likelihood functions. Other choices of pseudo R^2 include the specifications of Cragg-Uhler and Chow (Daykin and Moffat, 2002; Greene 2000, p.683).

- For undergraduate level 2 (UG2), all but two of the ten independent variables are statistically significant. Furthermore, their magnitudes show that they have substantial impact on *TEVAL*. The coefficients of *RESPECT* and *ENTHUSM* are neither statistically significant nor are they numerically substantive.
- For UG3 seven out of the ten independent variables appear to be statistically significant (the exceptions being *RESPECT*, *ENTHUSM* and *THINKMEM*).
- For postgraduate (PG) estimated coefficients of all but three independent variables (*RESPECT*, *KNOWWELL* and *CONSULT)* appear significant. The negative sign of the coefficient of *RESPECT* appears to be counter-intuitive.

Based on the estimated equations in Table 2, the six most important factors that impact on *TEVAL* can be identified in order of the magnitudes of their coefficients and are set out in Table 3. Of these six factors the four that are common to all samples are: *EXPLAIN*, *ORGANIZE*, *PRESENT* and *LSKILLS*. However, their rankings, based on numerical magnitudes, vary across samples. For example, *EXPLAIN* and *PRESENT* are the two most important factors for UG2 sample followed closely by *ORGANIZE* and *LSKILLS*. For the UG3 sample *ORGANIZE* is by far the most important factor, while the next two factors *EXPLAIN* and *PRESENT* are close to each other followed by *LSKILLS* which is the distant fourth. For the PG sample, *ORGANIZE* is the most important attribute with *LSKILLS* not far behind. *PRESENT* and *EXPLAIN* exert relatively smaller influence on *TEVAL* for the PG sample.

Table 2: Ordered Probit Analysis of Overall Perceived Teaching Effectiveness Score (*TEVAL*) by Perceived Instructor Attributes

Variables	**Level 2 Undergraduate (UG2)**	**Level 3 Undergraduate (UG3)**	**Postgraduate (PG)**
CONSTANT	***4.592 (0.000)	***5.297 (0.000)	***3.040 (0.000)
ORGANIZE	***0.392 (0.000)	***0.660 (0.000)	***0.458 (0.000)
PRESENT	***0.424 (0.000)	***0.411 (0.000)	***0.251 (0.000)
FEEDBACK	0.040 (0.469)	***0.235 (0.003)	***0.191 (0.002)
RESPECT	**0.126 (0.037)	0.147 (0.152)	-0.050(.512)
KNOWWELL	***0.260 (0.000)	*0.240 (0.074)	0.100 (0.292)
ENTHUSM	0.062 (0.313)	0.113 (0.286)	**0.187 (.034)
THINKMEM	**0.120 (.043)	0.092 (0.353)	**0.171 (0.022)
EXPLAIN	***0.470 (0.000)	***0.457 (0.000)	***0.195 (0.002)
CONSULT	***0.145 (0.006)	***0.240 (0.004)	0.013 (0.846)
LSKILLS	***0.375 (0.000)	***0.268 (.003)	***0.378 (0.000)
μ_1	***1.451 (0.000)	***2.281 (0.000)	***1.450 (0.000)
μ_2	***3.770 (0.000)	***4.683 (0.000)	***3.380 (0.000)
μ_3	***6.034 (0.000)	***7.113 (0.000)	***5.222n (0.000)
$\chi^2(10)$	1218.52	588.16	732.24
N	929	490	823
Psuedo R^2	0.48	0.51	0.38

Notes: *p*-values (two-tail) are in parentheses. ***, **, and * respectively represent 1%, 5% and 10% significant levels for a two-tailed test.

It is also important to note that the four most important perceived attributes identified have much wider ranges in case of the UG3 sample compared to the UG2 and the PG samples. Furthermore, *KNOWWELL* is perceived to be a substantive factor for the level 2 undergraduate sample *ENTHUSM* is only an important factor in the postgraduate sample but not in other cases. *FEEDBACK* is an important factor for upper level courses while *CONSULT* is an important factor for both the undergraduate samples.

One disturbing, if surprising, feature of these results is that a key factor, *THINKMEM*, is statistically significant in only two out of three samples (UG2 and PG). More critically, in terms of numerical significance, it ranks 7th and 8th respectively for the PG and UG2 sample in the list of ten instruction attributes.

Table 3: Six Most Important Factors Influencing *TEVAL* in Order of the Magnitudes of Their Coefficients by Level and Program

Ranking	**Level 2 Undergraduate**	**Level 3 Undergraduate**	**Postgraduate**
1 (Highest)	*EXPLAIN* (0.470)	*ORGANIZE* (0.660)	*ORGANIZE* (0.458)
2	*PRESENT* (0.424)	*EXPLAIN* (0.457)	*LSKILLS* (0.378)
3	*ORGANIZE* (0.392)	*PRESENT* (0.411)	*PRESENT* (0.251)
4	*LSKILLS* (0.375)	*LSKILLS* (0.268)	*EXPLAIN* (0.195)
5	*KNOWWELL* (0.260)	*CONSULT* (0.240)	*FEEDBACK* (0.191)
6 (Lowest)	*CONSULT* (0.145)	*FEEDBACK* (0.235)	*ENTHUSM* (0.187)
Range	0.325	0.425	0.271

These results seem consistent with some previous studies such as that of Boex (2000) who found that organization and clarity is the most important perceived instruction attribute influencing the overall teaching effectiveness score. However, Boex relied on highly aggregated data and it was not clear if the impact of these factors differed across levels and programs. The present study may seem to represent a replication of Boex's study with Australian data. However, the main difference between the two studies is that in the present paper there was no apparent need to engage in factor analysis due to the more succinct nature of the SET instrument.

4 SOME FURTHER ANALYSIS OF EMPIRICAL RESULTS

In light of the preceding discussion, this section predicts the probability of *TEVAL* score when perceived attributes used for the estimated equations in Table 2 show an increase or a decrease. We start with a base case where all attributes are given a rating of 4 and estimate the corresponding probability of an instruction getting a student rating of 5. It can be seen from Table 4 that in the base case, the estimated probability of the instruction being rated 5 is appreciably higher in case of the PG sample than in either of the two UG samples. On the other hand an instruction with all perceived attributes set at 4, has just over a 16 per cent chance of a rating of 5 for the UG sample while the base case yields nearly 25 per cent chance of a student rating of 5 for the PG sample.

Let us now consider two alternative scenarios in order to predict probability of an instruction getting a 5 when ratings of all attributes are (1) increased from 4 to 5; and (2) decreased from 4 to 3.

The results of the exercise based on the first scenario are set out in Table 4. The probability of getting a 5 for teaching effectiveness is most influenced by *ORGANIZE, EXPLAIN, PRESENT, and LSKILLS.*[12] The degree of variation differs across levels and programs. For example:

- In case of UG2, increasing the score of *ORGANIZE* from 4 to 5, *paribus* increases the probability of *TEVAL* = 5 from 16.6 to 28.2 per cent. The respective marginal effects of increasing the scores from 4 to 5 *ceteris paribus* in *EXPLAIN, PRESENT* and *LSKILLS* lead to the increases in the probabilities of 30.9, 29.3 and 27.6 per cent in a *TEVAL* score from 4 to 5 from the base level of 16.6 per cent.

[12] Predicted probabilities for the remaining six perceived attributes were estimated but not reported in Table 4 and Table 5 for brevity. However, they can be made available upon request.

- For UG3, increasing *ORGANIZE* from 4 to 5 the probability of *TEVAL* is likely to increase from 16.9 to 38.3 per cent while the same margin of change in *PRESENT*, *EXPLAIN*, and *LSKILLS* is likely to increase the probability of *TEVAL* respectively to 29.3, 30.9 and 24.5 per cent.
- For the PG sample a transition from 4 to 5 in respect of *ORGANIZE* is likely to increase the probability of a *TEVAL* score of 5 from a base level 24.7 per cent to 41.1 per cent. A transition form 4 to 5 in respect of *PRESENT*, *EXPLAIN*, and *LSKILLS* is likely to increase the probability of *TEVAL* score of 5 respectively to 33.3, 31.3 and 38.0 per cent from the same base level of 24.7 per cent.

Table 4: Variations in Predicted Probability (Measured in Percentage) of *TEVAL* Due to Increases in the Rating of Four Most Influential Attributes from 4 to 5 *ceteris paribus*.

Level 2 Undergraduate

Probability	Base case*	*ORGANIZE*	*PRESENT*	*EXPLAIN*	*LSKILLS*
TEVAL=4	73.6	67.2	66.5	65.2	67.6
TEVAL=5	16.6	28.2	29.3	30.9	27.6

Level 3 Undergraduate

Probability	Base case*	*ORGANIZE*	*PRESENT*	*EXPLAIN*	*LSKILLS*
TEVAL=4	76.0	60.0	67.7	66.5	71.4
TEVAL=5	16.9	38.3	29.3	30.9	24.5

Postgraduate

Probability	Base case*	*ORGANIZE*	*PRESENT*	*EXPLAIN*	*LSKILLS*
TEVAL=4	63.0	53.6	58.8	59.9	55.8
TEVAL=5	24.7	41.1	33.3	31.3	38.0

*All attributes =4.

Likewise one can observe (from Table 5) varying types and degrees of variations in predicted probabilities of *TEVAL* to a transition from a rating of 4 to a rating of 3 in respect of the four attributes considered above. For instance, the predicted probability of *TEVAL* score of 5 is most adversely affected by a decline from 4 to 3 in the rating of:

- *EXPLAIN* (from 16.3 per cent to 7.5 per cent) followed closely by *PRESENT* (8.2 per cent), *ORGANIZE* (8.7 per cent) and *LSKILLS* (8.9 per cent) for the UG2 sample.
- *ORGANIZE* (per cent to 5.3 per cent) followed closely by *EXPLAIN* (7.9 per cent) and *PRESENT* (8.6 per cent) and somewhat distantly by *LSKILLS* (11.0 per cent) for the UG3 courses from 16.9 per cent for the UG3 sample.
- *ORGANIZE* (12.7 per cent) followed closely by *LSKILLS* (14.4) and somewhat distantly by *PRESENT* (17.5 per cent) and *EXPLAIN* (19 per cent) from 24.7 per cent for the PG sample.

Table 5: Variations in Predicted Probability (Measured in Percentage) of *TEVAL* Due to Decreases in the Rating of Four Most Influential Attributes from 4 to 3 *ceteris paribus*

Level 2 Undergraduate

Probability	Base case*	*ORGANIZE*	*PRESENT*	*EXPLAIN*	*LSKILLS*
TEVAL=4	73.6	73.0	72.6	72.0	73.2
TEVAL=5	16.6	8.7	8.2	7.5	8.9

Level 3 Undergraduate

Probability	Base case*	*ORGANIZE*	*PRESENT*	*EXPLAIN*	*LSKILLS*
TEVAL=4	76.0	73.9	77.0	76.7	77.6
TEVAL=5	16.9	5.3	8.6	7.9	11.0

Postgraduate

Probability	Base case*	*ORGANIZE*	*PRESENT*	*EXPLAIN*	*LSKILLS*
TEVAL=4	63.0	63.2	64.3	64.3	63.8
TEVAL=5	24.7	12.7	17.5	19.0	14.4

* All attributes =4

5 CONCLUDING OBSERVATIONS

Employing SET data and ordered probit analysis this paper finds that instructor's improvement in organization, presentation, explanation, improving students' learning skills, significantly positively and substantively impact on students' perceptions of teaching effectiveness when *TEVAL* scores are used to measure this effectiveness. The converse also appears to hold. The impacts of these factors vary between postgraduate and lower undergraduate courses as well as between levels within the undergraduate program. Furthermore, it was found that scores tend to be systematically influenced by whether the subject is at a lower level or not.

A particular feature of this paper is that results provide a range of useful information which can help improve *TEVAL* score. For example, lower undergraduate students most prefer an instruction with clear explanation and presentation of materials in an interesting way. On the other hand, students at the upper undergraduate and postgraduate levels most prefer instruction involving lecture classes that are perceived to be well organized, presented and clearly explained and that help them in improving their learning skills.

An instructor looking at his or her *TEVAL* results, in conjunction with the empirical findings of this paper, should have a strong message about how to improve his/her *TEVAL* score, and a department chair would gain a strong impression of the teaching strengths and weaknesses of an instructor. For example, the department chair could note that although an instructor has a high *TEVAL* score, s/he has a low *THINKMEM* (emphasizing thinking rather than memorizing) score. Alternatively, an instructor with a low *TEVAL* could have demonstrate strength in *THINKMEM* and improving learning skills of students both of which are extremely important pedagogical responsibilities of an instructor.

One disturbing and somewhat surprising finding of this paper is that, even though *THINKMEM* is a statistically significant factor in two of the three samples (UG2 and PG), their numerical influences are not substantive at any level of instructions. In our view that should raise some alarm in the sense that high *TEVALs* can be achieved at the cost of some critically important factors in teaching and learning.[13]

In conclusion, given the limitations of the data stated in Section 2, the findings of this paper are subject to the following *caveats*:

- Omission of a number of relevant factors encompassing student and instructor attributes is likely to result in incomplete specification of the model and can have serious econometric consequences (see, for example. Deegan, 1976, pp.237-38).
- Given that *TEVAL* score is more likely to be a subjective measure than an objective one, gives rise to the measurement errors in variables (Judge et al. 1988, p.582; Griliches 1974, pp.973-74).
- A number of independent variables that determine SET occur simultaneously and, therefore, the problem of simultaneity bias is likely to be present (Krautman and Sander 1999).
- For those who want to follow a general-to-specific methodology in statistical/econometric analysis (Gerrard 1995), the available database still remains an obstacle to its implementation.

[13] One needs to be reminded though that the ordered probit analysis itself suffers from limitations. One such limitation, for example, is that the relationship it detects in relation to each of its component variables is only monotonic. Furthermore, it involves an additive function but a multiplicative relationship can sometimes be important. The nature of the mathematical relationship affects the possible results (associations) obtained. For example, *THINKME*M could improve *TEVAL* scores up to a point but the cause these to decline if the instructor makes the students think too much. But the general relationship may be positive. Again the scaling of instructor attributes such as *THINKMEM* can vary across students because how these are to be determined is left open.

ACKNOWLEDGEMENTS

The authors wish to thank Chris O'Donnell for assistance with the econometric modeling and Asad Khan for extensive consultation with the statistical aspects of this paper. We gratefully acknowledge Nghiem Hong Son's help with excellent research assistance and our (anonymous) colleagues who kindly provided access to their *TEVAL* data for this research. Usual *caveats* apply.

REFERENCES

Abrami, P.C. (1989) "How Should We Use Student Ratings to Evaluate Teaching?", *Research in Higher Education*, **30**(2), pp.221-27.

Akerloff, G. (1970) "The Market for Lemons: Quality Uncertainty and the Market Mechanism", *Quarterly Journal of Economics*, **84**, pp.488-500.

Alauddin, M. and Butler, J.E. (2004a) "From a Vicious Circle of Anxiety to a Virtuous Circle of Learning: Experience of Teaching Statistics to a Heterogeneous Clientele", *American Journal of Applied Sciences*, **1**(3), pp.202-208.

Alauddin, M. and Butler, J.E. (2004b) "Teaching Economics in a Changing University Environment: Some Australian Experience", *International Journal of Social Economics*, **31** (7-8), pp.706-20.

Alauddin, M. and Tisdell, C.A. (2000) "Changing Academic Environment and Teaching of Economics at the University-level: Some Critical Issues Analysed with the Help of Microeconomics", *Economic Papers*, **19**(1), pp.1-17.

Aleamoni, L.M. (1999) "Student Rating Myths Versus Research Facts from 1924 to 1998", *Journal of Personnel Evaluation in Education*, **13**(2), pp.153-66.

Arreola, R.A. (1995) *Developing A Comprehensive Faculty Evaluation System*, Bolton, MA: Anker.

Becker, W.E. (2000) "Teaching Economics in the 21st Century", *Journal of Economic Perspectives*, **14**(1), pp.109-19.

Boex, L.F.J. (2000) "Attributes of Effective Economic Instructors: An Analysis of Student Evaluations", *Journal of Economic Education*, **31**(3), pp.211-27.

Centra, J.A. (1993) *Reflective Faculty Evaluation; Enhancing Teaching and Determining Faculty Evaluation*, San Francisco: Jossey-Bass.

Chan, G., Miller, P.W. and Tcha, M.J. (2005) "Happiness in University Education", *International Review of Economics Education*, **4**(1), pp.20-45.

D'Apollonia, S. and Abrami, P.C. (1997) "Navigating Student Ratings of Instruction", *American Psychologist*, **52** (11), pp.1198-1208.

Daykin, A. R. and Moffatt, P. G. (2002). "Analyzing Ordered Responses: A Review of the Ordered Probit Model", *Understanding Statistics, 1*(3), 157.

DeCanio, S.J. (1986) "Student Evaluations of Teaching: A Multinomial Logit Approach", *Journal of Economic Education*, **17**(3), pp.165-76.

Deegan Jr., J. (1976) "The Consequences of Model Missepcification in Regression Analysis", *Multivariate Behavioral Research*, **11**(2), pp.237-48.

Felton, J., Mitchell, J. and Stinson, M. (2004) "Web-Based Student Evaluations of Professors: The Relations Between Perceived Quality, Easiness and Sexiness", *Assessment and Evaluation in Higher Education*, **29**(1), pp.91-108.

Gerrard, B. (1995) "The Scientific Basis of Economics: A Review of Methodological Debates in Economics and Econometrics", *Scottish Journal of Political Economy*, **42**(2), pp.221-34.

Greene, W.H. (2000) *Econometric Analysis*, New Jersey: Prentice-Hall.

Greenwald, A.G. and Gillmore, G.M. (1997) "Grading Leniency is a Removable Contaminant of Student Ratings", *American Psychologist*, **52** (11), pp.1209-17.

Griliches, Z. (1974) "Errors in Variables and Other Unobservables", *Econometrica*, **42**(6), pp.971-98.

Judge, G.G., Hill, R.C., Griffiths, W.E., Lütkepohl, H. and Lee, T-C. (1988) *Introduction to the Theory and Practice of Econometrics*, New York: John Wiley.

Krautmann, A. and Sander, W. (1999) “Grades and Student Evaluation of Teachers”, *Economics of Education Review*, **18**(1), pp.49-53.

Laudan, L. and Leplin, J. (1991) “Empirical Equivalence and Underdetermination”, *Journal of Philosophy*, **88**, pp.449-72.

Marsh, H. (1987) “Students’ Evaluation of University Teaching: Research Findings, Methodological Issues, and Directions for Future Research”, *International Journal of Educational Research*, **11**(3), pp.263-388.

Marsh, H. and Roche, L.A. (1997) “Making Students’ Evaluation of Teaching Effectiveness Effective: The Critical Issues of Validity, Bias and Utility”, *American Psychologist*, **52**(11), pp.1187-97.

Mason, P. M., Steagall, J. W. and Fabritius, M. M. (1995) “Student Evaluation of Faculty: A new Procedure for Using Aggregate Measures of Performance”, *Economics of Education Review*, **14** (4), pp. 403-416

McKlvey, R.D. and Zavoina, W. (1975) “A Statistical Model for the Analysis of Ordinal Level Dependent Variables”, *Journal of Mathematical Sociology*, **4**(1), pp.103-20.

Sproule, R. (2002) “The Underdetermination of Instructor Performance by Data from the Student Evaluation of Teaching”, *Economics of Education Review*, **21**, pp. 287-294.

Tang, T. L-P. (1997) “Teaching Evaluation at a Public Institution of Higher Education: Factors Related to Overall Teaching Effectiveness”, *Public Personnel Management*, **26**(3), pp.379-89.

Wachtel, H.K. (1998) “Student Evaluation of College Teaching Effectiveness: A Brief Review” *Assessment and Evaluation in Higher Education*, **23**(2), pp.191-211.

Wilson, R. (1998) “New Research Casts Doubt on Value of Student Evaluations of Professors”, *Chronicle of Higher Education*, (January 16), p.A12.

Wright, R.E. (2006) “Student Evaluations of Faculty: Concerns Raised in the Literature, and Possible Solutions”, College Student Journal, **40**(2), pp.417-22.